

# Natural images from the birthplace of the human eye

Gašper Tkačik<sup>1,2,\*</sup>, Patrick Garrigan<sup>3</sup>, Charles Ratliff<sup>4</sup>, Grega Milčinski<sup>5</sup>, Jennifer M. Klein<sup>7</sup>, Lucia H. Seyfarth<sup>6</sup>, Peter Sterling<sup>6</sup>, David Brainard<sup>7</sup>, and Vijay Balasubramanian<sup>1</sup>

<sup>1</sup> *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

<sup>2</sup> *Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria*

<sup>3</sup> *Department of Psychology, Saint Joseph's University, Philadelphia, PA 19131, USA*

<sup>4</sup> *Department of Ophthalmology, Northwestern University, Chicago, IL 60611, USA*

<sup>5</sup> *Sinergise d.o.o, SI-1000 Ljubljana, Slovenia*

<sup>6</sup> *Department of Neuroscience, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA*

<sup>7</sup> *Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA*

\* *E-mail: gtkacik@ist.ac.at*

Here we introduce a database of calibrated natural images publicly available through an easy-to-use web interface. Using a Nikon D70 digital SLR camera, we acquired about 5000 six-megapixel images of Okavango Delta of Botswana, a tropical savanna habitat similar to where the human eye is thought to have evolved. Some sequences of images were captured unsystematically while following a baboon troop, while others were designed to vary a single parameter such as aperture, object distance, time of day or position on the horizon. Images are available in the raw RGB format and in grayscale. Images are also available in units relevant to the physiology of human cone photoreceptors, where pixel values represent the expected number of photoisomerizations per second for cones sensitive to long (L), medium (M) and short (S) wavelengths. This database is distributed under a Creative Commons Attribution-Noncommercial Unported license to facilitate research in computer vision, psychophysics of perception, and visual neuroscience.

## INTRODUCTION

High-resolution digital cameras are now ubiquitous and affordable, and are increasingly incorporated into portable computers, mobile phones and other handheld devices. This accessibility has led to an explosion of online image databases, accessible through photo-sharing websites such as Flickr and Smug-Mug, social networks such as Facebook, and many other internet sites. Disciplines such as neuroscience, computer science, engineering and psychology have profited from research into the statistical properties of natural image ensembles [1, 2], and it might seem that the side benefit of public photo-sharing websites is an ample supply of image data for such research. Some research, however, requires carefully calibrated images that accurately represent the light that reaches the camera sensor. For example, to address questions about the early visual system, images should not deviate systematically from the patterns of light incident onto, and encoded by, the retina. Specific limitations of uncontrolled databases that hinder their use in vision research include **(i)** compression by lossy algorithms, which distort image structure at fine scales; **(ii)** photography with different cameras, which results in unpredictable quality and incomparable pixel values; **(iii)** photography with different lenses or focal lengths, which may introduce unknown optical properties to the image incident on the sensor, and **(iv)** photography for unspecified purposes, which may bias image content toward faces, man-made structures, panoramic landscapes, etc., while under-representing sky, ground, feces, or other more mundane content.

Due to these limitations, research involving natural images typically relies on a well-calibrated database [3–5]. Examples where good image databases are essential include early visual processing in neural systems [4, 6–11], computer vision algorithms [12–14], and image compression/reconstruction methods [15, 16]. In each of these fields, much of the research requires accurate characterization of the statistics of natural image ensembles. Natural images exhibit characteristic luminance distributions [17] and spatial correlations [7, 18]. These result from environmental regularities such as physical laws of projection and image formation, natural light sources, and the reflective properties of natural objects. Further, natural images exhibit higher-order regularities such as edges, shapes, contours and textures

that are perceptually salient, but difficult to quantify [19–24]. To characterize these regularities in image patches of increasing spatial extent, the number of image samples required to collect reliable joint pixel luminance probability distributions scales exponentially with the number of pixels in the patch, ultimately limiting the reliability of our estimates. Nevertheless, with sizable databases one might push the sampling limit to patches of up to  $\sim 10 \times 10$  pixels discretized into a few luminance levels, and even at that restricted size interesting results have emerged [25–27].

In this paper we describe a collection of calibrated natural images, which tries to address the shortcomings of uncontrolled image databases, while still providing substantial sampling power for image ensemble research. The database is organized into themed sub-collections and is broadly annotated with keywords and tags. The raw images are linear with incident light intensity in each of the camera’s three color channels (Red, Green, and Blue). Further, because much research is concerned with human vision, we have translated each image into physiological units relevant to human cone photoreceptors. The raw image data, demosaiced RGB images, grayscale images representing luminance, and images in the physiologically relevant cone representation are available in the database.

## RESULTS

### Image calibration

We checked **(i)** that the Nikon D70 camera has a consistent resolution (in pixels per degree) in the vertical and horizontal directions; **(ii)** that its sensor responses are linear across a large range of incident light intensities; and **(iii)** that its shutter, aperture, and ISO settings behave regularly, so that it is possible to estimate incident photon flux given these settings. We **(iv)** measured the spectral response of the camera’s R, G, and B sensors by taking images of a reflectance standard illuminated by a series of 31 narrowband light sources and verified that these measured responses allowed us to predict the R, G, and B responses to spectrally broadband light. We also **(v)** characterized the “dark response” of the camera, i.e. the sensor output with no light incident onto the CCD (charge-coupled device) chip. Finally, we **(vi)** measured the spatial modulation transfer function of each of the R, G, and B sensor planes for broadband incident light. We examined two separate D70 cameras, and found that they yielded consistent results and could be used interchangeably, except for a single constant scaling factor in the CCD sensor response. Taken together, the camera measurements allow us to transform raw camera RGB values into standardized RGB response values that are proportional to incident photon fluxes seen by the three camera sensor types. Details on the camera characterization are provided in *Materials and Methods: Camera response properties*.

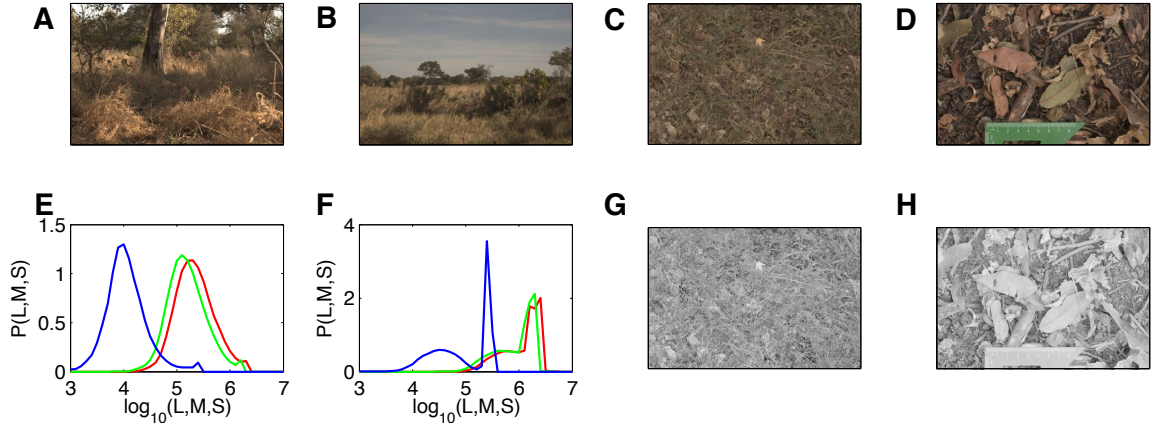
With these results in hand, we can further transform the standardized RGB data into physiologically relevant quantities (see details in *Materials and Methods: Colorimetry*). The photon flux arriving at the human eye is filtered by passage through the ocular media before entering a cone photoreceptor aperture. There, with some probability, each photon may cause an isomerization. There are three classes of cones, the L, M, and S cones. Each contains a different photopigment, with the spectral sensitivities of the three photopigments peaking at approximately  $\lambda_S = 421$  nm,  $\lambda_M = 530$  nm, and  $\lambda_L = 559$  nm, respectively [28]. We have transformed the RGB images into physiologically-relevant estimates of the isomerization rates that the incident light would produce in human L, M, and S cones. For this transformation, we used the Stockman-Sharpe/CIE 2-degree (foveal) estimates of the cone fundamentals [28, 29], together with methods outlined by Yin et al. [30] to convert nominal cone coordinates to estimates of photopigment isomerization rates. We have also transformed the RGB images into grayscale images representing estimates of the luminance of the incident light (in units of candelas per meter squared with respect to the CIE 2007 two-degree specification for photopic luminance spectral sensitivity). An important contribution of this work is to report image data from a biologically relevant environment in

units of cone photopigment isomerizations, thus characterizing the information available at the first stage of the human visual processing pathways.

### Image ensemble

We have assembled a large database of natural images acquired with a calibrated camera (see Materials and Methods), and made it available to the general public through an easy-to-use web interface. The extensive dataset covers a single environment: a rich riverine / savanna habitat in the Okavango delta, Botswana, which is home to the full panoply of vertebrate and invertebrate species, such as lion, leopard, cheetah, elephant, warthog, antelope, zebra, giraffe, various bird species etc. This environment was chosen because it is thought to be similar to the environment where the human eye, and retina in particular, have evolved. The database consists of  $\sim 5000$  images, organized into about 100 folders (albums).

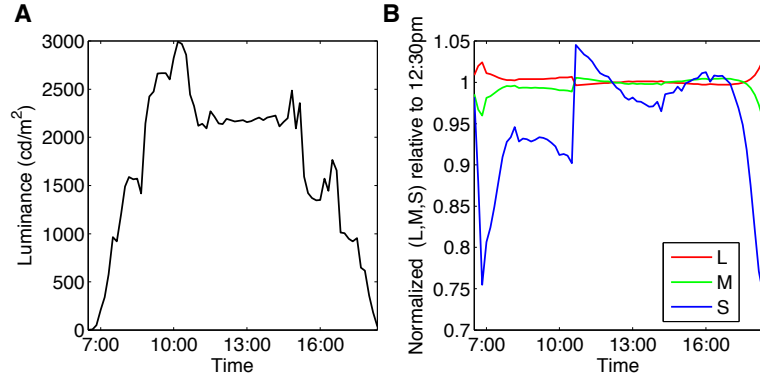
Tables I and II provide an album-by-album summary of the database content, along with content keywords for each album and tags that give additional information about how the images were gathered. Figure 1 shows several examples of images from the database: baboon habitat, panoramic images that include the horizon, closeups of the ground, and closeup images that include the ruler for absolute scale determination.



**FIG. 1. Example images from the Botswana dataset.** **A-D)** Some natural scenes from various albums, including a tree, grass and bushes environment, the horizon with a large amount of sky, and closeups of the ground; the last image is from the image set containing a ruler than can be used to infer the absolute scale of objects. **E-F)** The distributions of L (red), M (green) and S (blue) channel intensities across the image for images A) and B), respectively. The large sky coverage in B) causes a peak in the S channel at high values. The horizontal axis is log base 10 of pigment photoisomerization rates per second. **G-H)** Grayscale images showing log luminance corresponding to the images in C) and D), respectively.

Figure 2 shows a simple analysis of images from album CD32B. Images of cloudless sky were taken every 10 minutes, from 6:30 until 18:30. We compute the average luminance as a function of time, and the LMS color composition of the light incident on the camera that was pointing up at the homogenous sky.

Figure 3 shows an analysis of 23 images from album CD04B, where closeups of the grass scrub on the ground were taken from various distances. We computed the pairwise luminance correlation function as a function of pixel separation, and show how it varies systematically with distance to the ground.



**FIG. 2. The color content and luminance of the sky.** **A)** The luminance in candelas per square meter shows the rise in the morning and decay in the evening, along with the fluctuations during the day. **B)** The color content of the sky. To report relative changes in color content corrected for overall luminance variation, the L, M and S channels have been divided by their reference values at 12:30pm to bring the three separate curves together at the 12:30pm time point. In addition, all three curves were multiplied by the luminance at 12:30pm and then divided by the luminance at the time each measurement was taken. At sunrise and sunset, the L (redder) channel is relatively more prominent.

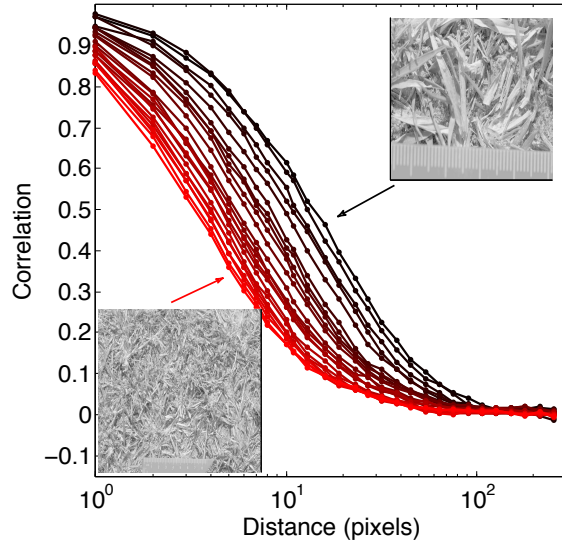
Since the grass scrub on our images has a preferred scale, the correlation function should decay faster the farther away from the scene the camera is, and this is indeed what we see. Scale invariance in natural scenes presumably emerges because of the distribution of object sizes and distances from which the objects are viewed [31]. Our image collection can be used to study the scaling properties of the ensemble systematically.

The two examples provided above are intended to illustrate the strengths of our database: calibration and normalization into physiologically relevant units; organization into thematic (keywords) and methodological sequences (tags), that explore variations between the scenery content and variations induced by systematic changes in controllable parameters; and a thorough sampling that should suffice for the estimation of higher-order statistics or even accumulation of image patch probability distributions.

In assembling the database, we could have taken an alternative approach and built a catalogue of as many objects from the environment as possible, photographed at a chosen “standard” camera position and controlled illumination, removed from the natural context and placed on a neutral background. While there are advantages to exhaustively pursuing this approach, we regarded it as focusing on natural *objects* and not on natural *scenes*. Nevertheless we provide a limited set of albums in the collection of Table II, where closeups of fruit, grass, bark are provided; we reasoned that some of these closeups would be of use for studying properties of natural objects such as texture.

### Data access

The image database is accessible at <http://tofu.physics.upenn.edu/~upennidb>, or through anonymous FTP at <ftp://anonymous@tofu.physics.upenn.edu/fulldb>. A standard gallery program for viewing the images on the web makes browsing and downloading individual images or whole albums easy [32]. Once images have been added to the ‘cart’ and the user selects the download option, the database prompts the user to select the formats for download; the available formats are (i) NEF (raw camera sensor output, Nikon proprietary format); (ii) RGB Matlab matrix (demosaiced RGB values before dark

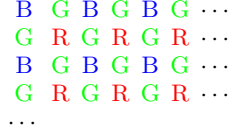


**FIG. 3. Pairwise correlations in natural scenes.** We analyzed 23 images of the same grass scrub scene, taken from different distances (black – smallest distance, red – largest distance). For every image, we computed the pixel-to-pixel correlation function in the luminance channel, and normalized all correlation functions to be 1 at  $R = 0$  pixels. For largest distances,  $R = 256$  pixels, the correlations decay to zero. The decay is faster in images taken from afar (redder lines, the largest distance image shown as an inset in the lower left corner), than in images taken close up (darker lines, the smallest distance image shown as an inset in the upper right corner). All images contain a green ruler that facilitates the absolute scale determination; for this analysis, we exclude the lower quarter of the image so that the region containing the ruler is not included in the sampling.

response subtraction); **(iii)** LMS Matlab matrix (physiological units of isomerizations per second in L, M, S cones); **(iv)** LUM Matlab matrix (grayscale image in units of  $\text{cd/m}^2$ ); and **(v)** AUX meta-data Matlab structure (containing camera settings and basic image statistics). *Materials and Methods: Image extraction and data formats* documents the detailed image processing pipeline. After image format selection has been made, the database prepares a download folder which contains the selected images in the requested formats and with the album (directory) structure maintained. The folder is available for download by (recursive) FTP, for instance by using an open-source WGET tool [33]; in this way large amounts of data, e.g., the whole database ( $\sim 300$  Gb of image data for all available formats combined) can be reliably transferred. A list of requested images is provided with the image files, and can be used to retrieve the same selection of images from the database directly; this facilitates the reproducibility of analyses and uniquely defines each dataset. All images are distributed under Creative Commons Attribution-Noncommercial Unported license [34], which prohibits commercial use but allows free use and information sharing / remixing as long as authorship is recognized. Images and processing software used to calibrate the cameras is available upon request.

## MATERIALS AND METHODS

**Camera response properties.** Two D70 (Nikon, Inc., Tokyo, Japan) cameras are described. Each was equipped with an AF-S DX Zoom-Nikkor 18-70mm f/3.5-4.5G IF-ED lens. To protect the front surface of the lens, a 52 mm DMC (Digital Multi Coated) UV skylight filter was used in all measure-



**FIG. 4.** A fragment of the mosaic pattern that tiles the CCD sensor. Each pixel is either red (R), green (G), or blue (B). The pixels are present in ratios 1:2:1 in the CCD array. The upper-left hand corner of the fragment matches the upper-left hand corner of the raw image data decoded by DCRAW.

ments. One camera, referred to in this document as the *standard camera* (serial number 2000a9a7), is characterized in detail. A subset of the measurements was made with a second D70 camera, referred to in this document as the *auxiliary camera* (serial number 20004b72). Its response properties matched those of the standard camera after all responses were multiplied by a single constant.

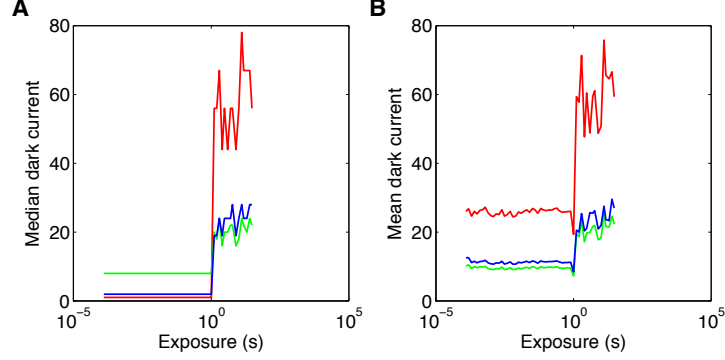
*Geometric information.* The D70’s sensor provides a raw resolution of 3040(h)  $\times$  2014(v) pixels. The angular resolution of the camera was established by acquiring an image of a meter stick at a distance of 123 cm. Both horizontal and vertical angular resolutions were 92 pixels per degree. This resolution is slightly lower than that of foveal cones, which sample the image at approximately 120 cones per degree [35].

*Image quantization.* The camera documentation indicates that the D70 has native 12-bit-per-pixel intensity resolution [36], and that the NEF compressed image format supports this resolution. Nonetheless, the D70 appears not to write the raw 12-bit data to the NEF file. Rather, some quantization/compression algorithm is applied which converts the image data from 12 bit to approximately 9.4 bit per pixel resolution[37]. The raw data extraction program, DCRAW[65] [38], appears to correct for any pixel-wise nonlinearity introduced by this processing, but it cannot, of course, recover the full 12-bit resolution. The pixel values in the file extracted by DCRAW range between 0 and 16384 for red and blue, and between 0 and 16380 for the green channel, and these are the values we use for the analysis. We do not have a direct estimate of the actual precision of this representation.

*Mosaic pattern and block averaging.* The D70 employs a mosaic photosensor array to provide RGB color images. That is, each pixel in a raw image corresponds either to an R, G, or B sensor, as shown in Fig 4. R, G, and B values can then be interpolated to each pixel location. This process is known as demosaicing. The images in the database were demosaiced by taking each 2 by 2 pixel block and averaging the sensors responses of each type with that block (one R sensor, 2 G sensors, and 1 B sensor). This produced demosaiced images of 1519(h)  $\times$  1007(v) pixels; these images were used in all measurements reported below.

*Dark subtraction.* Digital cameras typically respond with positive sensor values even when there is no light input (i.e. when an image is acquired with an opaque lens cap in place). This dark response can vary between color channels, with exposure duration, with ISO and with temperature. We did not systematically explore the effect of all these parameters, but we did collect dark images as a function of exposure time for ISO 400 in an indoor laboratory environment. Dark image exposures below 1s generated very small dark response, with the exception of  $< 0.1\%$  of “hot pixels” that had high (raw value  $> 200$ ) response. The median dark response below 1s exposure is less than 11 raw units for all three color channels. For dark image exposures above 1s, the dark response jumped to  $\sim 20$  for G and B channels to and  $\sim 56$  for the R channel. The dynamic range of the camera in each color channel is approximately 0 – 16k raw units. Dark response values were subtracted from image raw response values as a part of our image preprocessing: for all exposures below or equal to 1s, red dark response value was taken as 1, blue dark response value was taken as 2, and green dark response value was taken as 8 raw units (these values correspond to the mean over exposure durations of the median values across pixels for exposures of less than or equal to 1s); for exposures above 1s, the measured median dark response values across the image frame, computed separately for each exposure, were subtracted. Raw pixel values that

after dark subtraction yielded negative values were set to 0. Figure 5A shows the dark response that was used for subtraction. Figure 5B shows the mean response, excluding the “hot” pixels, for comparison. All measurements reported below are for RGB values after dark subtraction. The dark response table is available as a supplementary material to this paper.



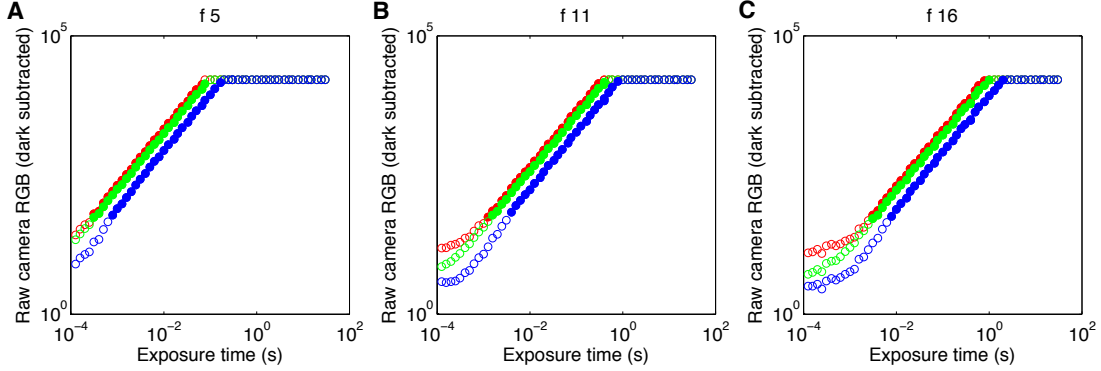
**FIG. 5. Dark response by color channel.** **A)** Dark response used for dark subtraction during image processing. For image exposure times below or equal to 1 s, the dark response for a given color channel (red, green, blue; plot colors correspond to the three color channels) was taken as the median over all the pixels of that color channel and over all dark image exposure times below 1 s; for image exposures above 1 s, we use the median over all the pixels of the same color channel at the given dark image exposure time. **B)** The mean value of dark response across all pixels of the image that are not “hot” (i.e. pixels with raw values  $< 200$ , more than 99% of pixels in each color channel), for each color channel, as a function of dark image exposure time. For all dark images, the camera was kept in a dark room with a lens cap on, with the aperture set to minimum ( $f22$ ), and ISO set to 400.

*Response linearity.* Fundamental to digital camera calibration is a full description of how image values obtained from the camera relate to the intensity of the light incident on the sensors. To measure the D70’s intensity-response function, a white test standard was placed 168 cm from the camera. The aperture was held constant at  $f5$ ,  $f11$  and  $f16$ . One picture was taken for each of the 55 possible exposure durations, which ranged from  $1/8000$  s to 30 s. Response values were obtained by extracting and averaging RGB values from a  $100 \times 100$  pixel region of the image corresponding to light from the white standard. The camera saturated in at least one channel for longer exposure durations. The camera responses were linear over the duration range  $1/320$  s to  $1/3$  s for the test light level and all apertures used, and for aperture  $f5$  approximate linearity extended down to smallest available exposure times, as shown in Fig. 6. Deviations from linearity are seen at short exposure durations for  $f11$  and  $f16$ . These presumably reflect a nonlinearity in the sensor intensity-response function at very low dark subtracted response values.

*ISO linearity.* To examine the effect of the ISO setting on the raw camera response, we acquired a series of 10 images of the white standard with the ISO setting changing from ISO 200 to ISO 1600 at exposures of  $1/125$  s and  $1/250$  s. In the regime where the sensors did not saturate, the camera response was linear in ISO for both exposure times, as shown in Fig. 7.

*Aperture test.* The aperture size (f-number) of a lens affects the amount of light reaching the camera’s sensors. For the same light source, the intensity per unit area for f-number  $x$ ,  $I(x)$ , relative to the intensity per unit area for f-number  $y$ ,  $I(y)$  should be given by:  $I(x) = (y/x)^2 I(y)$ . We tested this by measuring the sensor response as a function of exposure duration for all color channels at three different aperture sizes ( $f5$ ,  $f11$ , and  $f16$ ). We measured the sensor responses to the white test standard at various aperture sizes to confirm that the response was inversely proportional to the square of the f-number. The camera was positioned facing the white test standard illuminated by a slide projector with a tungsten





**FIG. 6. Linearity of the camera in exposure time.** The mean raw RGB response after dark subtraction of the three color channels (red, green, blue; shown in corresponding colors) is plotted against the exposure time in seconds. The values are extracted from images of a white test standard at  $f5$  (A),  $f11$  (B) and  $f16$  (C) and ISO 200 settings. Full plot symbols indicate raw dark subtracted values between 50 and 16100 raw units; these data points were used to fit linear slopes to each color channel and aperture separately. The fit slopes are 1.01 (R), 1.00 (G), 1.02 (B) for  $f5$ ; 1.00, 0.99, 1.02 for  $f11$ , and 1.01, 1.00, 1.02 for  $f16$ .

bulb in an otherwise dark room. The camera exposure time was held at  $1/250$  s and the ISO setting was 1000. Images of the white test standard (primary image region) were taken, one for each aperture setting between  $f1.8$  and  $f22$ . Because image values were saturated at the largest apertures ( $f1.8$  and  $f2$ ), we also extracted and analyzed image values from a less intense region in the same image series (secondary image region). The measurements, shown in Fig. 8A for the primary region and Fig. 8B for the secondary region, confirm that the aperture is operating correctly. There is some scatter of the measured points around the theoretical lines. This may be due to mechanical imprecision for each aperture. We did not pursue this effect in detail, nor attempt to correct for it. Nor did we investigate whether the slightly steeper slopes found for the blue channel represent a slight systematic deviation from expectations for the responses of that channel.

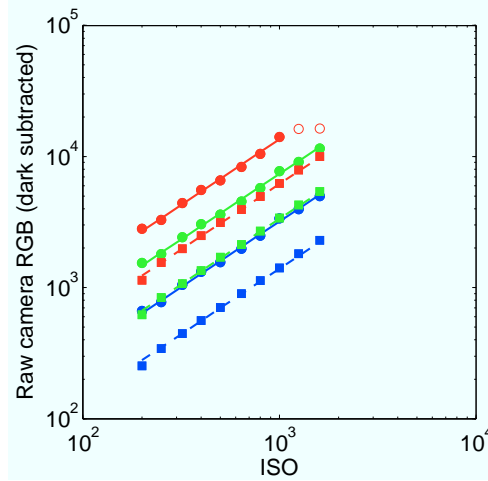
*Standardized RGB values.* After having verified that the sensor response scales linearly with ISO setting and the exposure, and as  $f^{-2}$  with the aperture number  $f$  across most of the camera's dynamic range, we define the *standardized RGB values* as dark-subtracted raw camera RGB values, scaled to the reference value of ISO 1000, reference aperture of  $f1.8$  and reference exposure time of 1 s:

$$\text{standardized\_RGB} = (\text{raw\_RGB} - \text{dark\_response}) \times \left( \frac{1000}{\text{ISO}} \right) \times \left( \frac{f}{1.8} \right)^2 \times \left( \frac{1\text{s}}{\text{exposure}} \right) \quad (1)$$

It is the standardized RGB values that provide estimates of the light incident on the camera across our image database. Our calibration measurements indicate that these values provide good estimates over the multiple decades of light intensity encountered in natural viewing. Scatter of measurements around the fit lines shown in the figures above does, however, introduce uncertainty of a few tenths of a log (base 10) unit in the intensity estimates on a finer intensity scale. Within image, optical factors are likely to introduce systematic variation in sensitivity from the center of the image to the edge, an effect that we did not characterize or correct for.

*Spectral response.* Next, we measured the spectral sensitivities of the camera sensors. A slide projector (Kodak Carousel 440 [39]), the Nikon D70 Camera, and a spectroradiometer [40] were positioned in front of the white test standard. Light from the projector was passed through one of 31 monochromatic



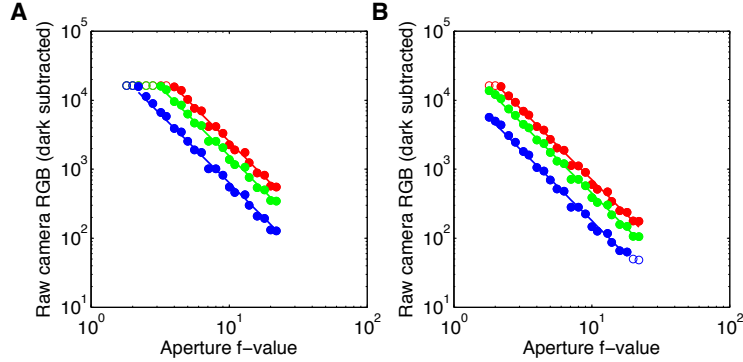


**FIG. 7. Linearity of the camera in ISO setting.** The mean raw RGB response of the three color channels (red, green, blue; shown in corresponding colors) is plotted against the ISO setting after dark subtraction, for two values of exposure time (solid line, circles = 1/125 s; dashed line, squares = 1/250 s), and  $f2.8$  aperture. The lines are linear regressions through non-saturated data points (solid squares or circles; raw dark subtracted values between 50 and 16100); the slopes are 0.99 (R), 0.98 (G), 0.99 (B) for 1/125 s exposure and 1.03, 1.02, 1.04 for 1/250 s exposure. The camera saturated in the red channel at longer exposure; the corresponding data points (empty red circles) are not included into the linear fit.

filters, which transmitted narrowband light between 400 nm and 700 nm, at 10 nm intervals. For each filter, a digital picture ( $f1.8$ , ISO 1000, and varying exposure duration) and a spectroradiometer reading were taken. The RGB data were dark subtracted and converted to standardized RGB values, and then compared to the radiant power read by the spectroradiometer to estimate the three sensor's spectral sensitivities. For this purpose, we summed power over wavelength and treated it as concentrated at the nominal center wavelength of each narrowband filter. For these images, dark subtraction was performed with dark images acquired at the same time as the spectral response images were acquired. These dark images were taken by occluding the light projected onto the white test standard, so that they accounted for any stray broadband light in the room as well as for sensor dark noise. The measured spectral sensitivities are shown in Fig. 9. These may be used to generate predictions of the camera sensor response to arbitrary spectral light sources. To compute predicted standardized RGB values, the input spectral radiance measured in units of  $\text{W}/(\text{nm} \cdot \text{sr} \cdot \text{m}^2)$  should be integrated against each spectral sensitivity.

To check the end-to-end accuracy of our camera RGB calibration, we acquired an image of the Macbeth color checker chart, extracted the raw RGB values for each of its 24 swatches, and converted these to the standardized RGB representation. We then compared these measured values against predictions obtained from direct measurements of the spectral radiance reflected from each swatch. The comparison in Fig. 10A shows an excellent fit between predicted and measured values.

*Spatial Modulation Transfer Function.* The modulation transfer function (MTF) describes how well contrast information is transmitted through an optical system. We estimated the MTF of our camera for the R, G, and B image planes individually, to compare the transmission of contrast information at various spatial frequencies for light intensity measured in each of the three types of camera photosensors. To do this, we first imaged a high-contrast black and white square-wave grating at a set of distances relative to the camera and extracted a horizontal grating patch from each image. This was summed over columns to

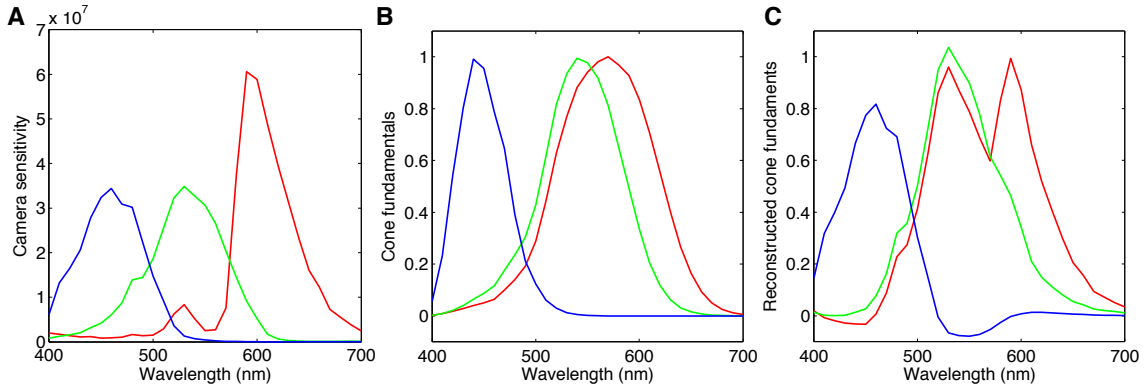


**FIG. 8. Camera response as a function of aperture size.** The raw dark subtracted response of the camera exposed to a white test standard (**A**) and a darker secondary image region (**B**), in three color channels (red, green, blue, shown in corresponding colors), as a function of the aperture (f-value), with exposure held constant to 1/250 s and ISO set to 1000. In the regime where the sensors are not saturated and responses are not very small (solid circles, raw dark subtracted values between 50 and 16100), the lines show a linear fit on a double logarithmic scale constrained to have a slope of  $-2$  (i.e.  $y = kx^{-2}$ ). Leaving the slopes as free fit parameters yields slopes of  $-2.02$  (R),  $-2.04$  (G),  $-2.10$  (B) for the primary region (white standard) in panel A, and  $-2.00$  (R),  $-2.03$  (G),  $-2.06$  (B) for the secondary region in panel B. Data points in the saturated or low response regime (empty circles) were not used in the fit.

The maximum absolute log base 10 deviation of the measurements from the fit lines is 0.1.

produce a one-dimensional signal for each image plane. We then used the fast-fourier transform to find the amplitude and spatial frequency of the fundamental component of the square wave-grating. In this procedure, we tried various croppings of the grating and chose the one that yielded maximum amplitude at the fundamental. This minimized spread of energy in the frequency domain caused by spatial sampling. We then reconstructed a one-dimensional spatial domain image by filtering out all frequencies except for the fundamental, and computed the image contrast as the difference between minimum and maximum intensity divided by the sum of minimum and maximum intensity. Figure 11 plots image contrast as a function of spatial frequency (cycles / pixel) for each color channel. At high spatial frequencies, the blue channel is blurred least and the red channel is blurred most, presumably due to chromatic aberration in the lens. We did not use these data in our image processing chain, but fits to the data are available (see caption of Fig. 11) either to correct for camera blurring for applications where that is desirable or to estimate the effect of camera blur on any particular image analysis. It should be noted that the MTF will depend on a number of factors, including f-stop, exactly how the image is focussed, and the exact spectral composition of the incident light. The measurements were made for  $f5.6$  and the camera's auto focus procedure. In addition, the measurements do not account for affects of lateral chromatic aberrations, which can produce magnification differences across the images seen by the three color channels. For these reasons, the MTF data should be viewed as an approximation.

*Camera comparisons.* We compared the response properties of two Nikon D70 cameras – the standard camera (used for the analysis detailed above), and a second auxiliary camera. We found that, under identical conditions, the sensor response of the auxiliary camera was approximately  $80\% \pm 5\%$  of the response of the standard camera, across all channels, response times, and apertures (see e.g. Fig. 12). We compared the two cameras imaging the white test standard with varying exposure durations, and for color swatches of the Macbeth color checker. We found that the ratio of camera responses was relatively constant throughout the linear range of exposure durations. To ensure that the response difference

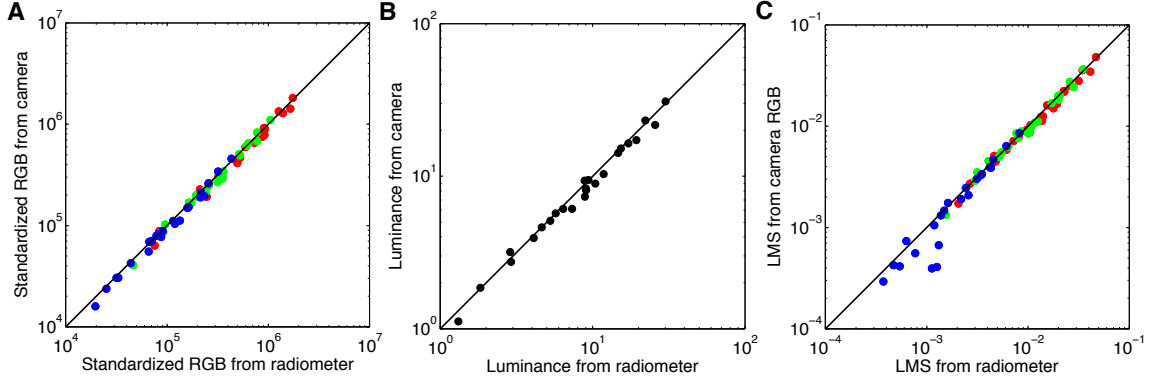


**FIG. 9. Spectral response of the camera.** **A)** The spectral sensitivity curves plotted here convert spectral radiance into standardized camera RGB values. **B)** The LMS cone fundamentals [29, 41] for L (red), M (green) and S (blue) cones. Note that the fundamentals are normalized to have a maximum of 1. **C)** A linear transformation can be found that transforms R,G,B readings from the camera with sensitivities plotted in (A) into reconstructed fundamentals L'M'S' shown here, such that L'M'S' fundamentals are as close as possible (in mean-squared-error sense) to the true LMS fundamentals shown in (B).

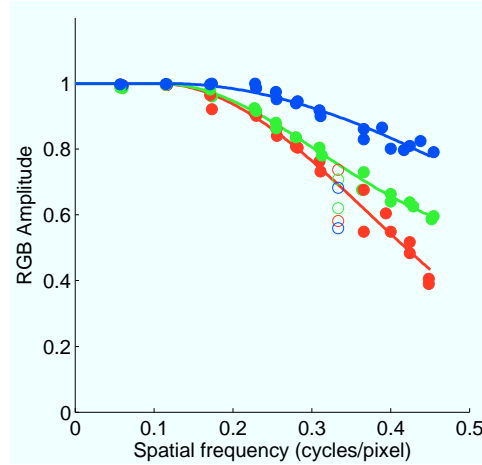
was not a result of the slightly different positions of the tripods on which each camera was mounted, we also took pictures of the white standard with the cameras on the same tripod (sequentially), then with the auxiliary camera 25 cm to the left of the standard camera and visa versa. The difference in camera responses could not be accounted for by small positional differences. Therefore, the different camera response magnitudes probably indicate a characteristic of the cameras themselves. At very fast or very slow exposure durations, the responses of both cameras were dominated by dark-response and sensor saturation, respectively, and the 80% response ratio did not apply. This analysis suggested that a complete calibration of the auxiliary camera was unnecessary. Instead, the spectral sensitivities of R, G, B sensors for the auxiliary camera were defined to be 0.80 times the spectral sensitivities of the standard camera; by using this simple multiplicative conversion to bring the auxiliary and standard cameras into accordance, the remaining image transformation steps remain unchanged between both cameras.

**Colorimetry.** For applications to human vision, it is useful to convert the camera RGB representation to one that characterizes how the incident light is encoded by the human visual system. For this reason, we used the camera data to estimate the photopigment isomerization rates of the human L, M, and S cones produced by the incident light at each pixel. Because the camera spectral sensitivities are not a linear transformation of the human cone spectral sensitivities, the estimates will necessarily contain some error [42], and different techniques for performing the estimation will perform better for some ensembles of input spectra than for others [43–45]. Here we employed a simple estimation method. We first found the linear transformation of the camera’s spectral sensitivities that provided the best least-squares approximation to the Stockman-Sharpe/CIE 2-degree (foveal) estimates of the cone fundamentals [29, 41]. Figure 9B plots the cone fundamentals and Fig. 9C plots the approximation to these fundamentals obtained from the camera spectral sensitivities. By applying this same transformation to the standardized RGB values in an image, we obtain estimates of the LMS cone coordinates of the incident light. For each cone class, these are proportional to the photopigment isomerization rate for that cone class.

To determine the constant of proportionality required to obtain LMS isomerization rates from the LMS cone coordinates for each cone class, we estimated photopigment isomerization rates directly from the measured spectral radiance of the light reflected from the 24 color swatches of the Munsell color

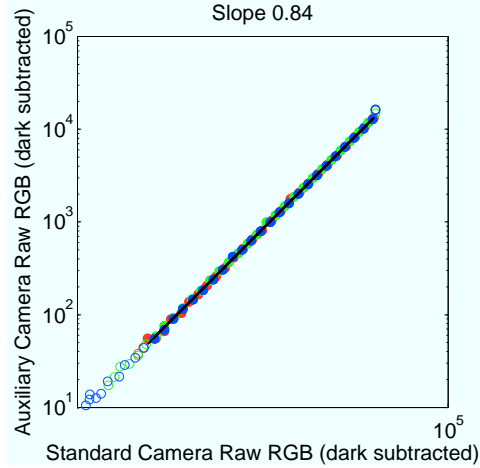


**FIG. 10. Checking the camera calibration.** Digital images and direct measurements of spectral radiance were obtained for the 24 color swatches of the Macbeth color checker chart. **A)** Raw standardized RGB values were obtained from the camera images as described in *Materials and Methods*. RGB response was also estimated directly from the radiometric readings via the camera spectral sensitivities shown in Fig. 9A. Plotted is the comparison of the corresponding  $24 \times 3$  RGB values; black line denotes equality. **B)** The luminance in  $\text{cd}/\text{m}^2$  measured directly by the radiometer compared to the luminance values obtained from the standardized camera RGB values. **C)** This plot shows the correspondence between the Stockman-Sharpe/CIE 2-degree LMS cone coordinates estimated from the camera and those obtained from the measured spectra. Plot symbols red, green and blue indicate L,M,S values respectively, and the data are for the 24 MCC squares.



**FIG. 11. Camera spatial MTF.** Estimated MTF is plotted as a function of spatial frequency for the red, green, and blue image planes (shown in corresponding colors). Solid lines show empirical fits to  $y = a \exp \{-b(x - d)^2\} + (1 - a) \exp(-cx^2)$ , where for all  $x \leq d$ ,  $y$  is set to 1 and where any fit values  $y$  greater than 1 were also set to 1. The fit parameters are  $a = 1.00, b = 6.96, c = 0.50, d = 0.10$  for red channel,  $a = 0.52, b = 12.37, c = 0.02, d = 0.11$  for green channel, and  $a = 1.00, b = 2.24, c = 0.50, d = 0.12$  for blue channel. MTF values at  $\sim 0.33$  cycles/pixel (empty plot symbols) systematically deviated from the rest and were excluded from the fit.

checker. This was done using software available as part of the Psychophysics Toolbox [30, 46] along with



**FIG. 12. Comparison between standard and auxiliary camera.** Raw dark subtracted values for three color channels (red, green, blue; shown in corresponding colors) of the same standard taken at different exposures using the standard and auxiliary cameras. Black line is a linear fit on non-saturated points (raw dark subtracted pixel values between 50 and 16100), with fit slope of 0.84. Images were acquired at  $f5.6$  and ISO 400. Similar slopes ( $0.8 \pm 0.05$ ) were found when positions of cameras imaging the white standard were slightly changed (see text), and when the camera readouts were compared on color swatches of the Macbeth color checker ( $f2.8$ , ISO 400.)

the parameter values for human vision provided in Table III. We also computed the response of the LMS cones from these same measured spectra using the cone response functions plotted in Fig. 9B. For each cone class, we then regressed the 24 isomerization rates against the 24 corresponding cone coordinates to obtain the scalar required to transform cone coordinate to isomerization rate. For the L, M, and S cones respectively, the scalar values are  $1.75 \cdot 10^5$ ,  $1.60 \cdot 10^5$ ,  $3.49 \cdot 10^4$ , with units that yield isomerizations per human cone per second.

For certain analyses, it is useful to provide grayscale versions of the images. We thus computed estimates of the luminance  $Y$  (in units of candelas per meter squared with respect to the CIE 2007 two-degree specification for photopic luminance spectral sensitivity [29]). These were obtained as a weighted sum of the estimated Stockman-Sharpe/CIE 2-degree LMS cone coordinates:  $Y = 433.94L + 275.82M - 0.09S$ ; these weights yield luminance in units of  $\text{cd/m}^2$ .

Figure 10B shows the comparison between the luminance for the Macbeth color checker the the luminance measured directly using the radiometer. Figure 10C shows the comparison between the Stockman-Sharpe/CIE 2-degree LMS cone coordinates estimated from the camera and those computed directly from the radiometrically measured spectra. Overall the agreement is good, with the exception of the S-cone coordinates for a few color checker squares. These deviations occur because the camera spectral sensitivities are not an exact linear transformation away from the cone fundamentals.

**Image extraction and data formats.** For each image in every database album (folder), the following operations are carried out:

1. **Raw image .NEF to .PPM conversion.** Images with name pattern `DSC_####.NEF` (where # are image serial numbers) were first extracted from the camera as proprietary Nikon Electronic Format (NEF) files (approximately 6Mb in size), which record “raw” sensor values. These files were converted to PPM files using `DCRAW v5.71` [38] in Document Mode (no color interpolation between RGB sensors) by using the `-d` flag, and written out as 48 bits-per-pixel (16 bits per color channel) PPM files by using the `-4` flag; we note that the behavior of `DCRAW` is highly version

dependent. In addition, we extracted the following image meta-data from the NEF file: the camera serial number (to determine whether an image was taken using the *standard* or *auxilliary* camera), exposure duration, f-value, and the ISO setting. We also extracted JPG formatted images of the NEF images; both JPG and NEF formats are available from the database as `DSC_####.NEF` and `DSC_####.JPG`.

2. **PPM to raw RGB conversion.** PPM images were loaded by our Matlab script as  $2014 \times 3038$  matrices. Because of the CCD mosaic, pixels in this matrix that have no corresponding CCD sensor contain zeros. Subsequently, each color plane of the image was block-averaged in  $2 \times 2$  blocks into a resulting  $1007 \times 1519 \times 3$  raw RGB format, where the last dimension indicates the color channel ( $R = 1, G = 2, B = 3$ ); red and blue pixels are weighted by 4 and green pixels by 2, reflecting the number of sensors for each color in every  $2 \times 2$  block. The resulting matrix has raw units spanning the range from 0 to 16384 for red and blue channels, and from 0 to 16380 for the green channel. It is available as a Matlab matrix with name pattern `DSC_####_RGB.mat` from the image database.
3. **Raw RGB to standardized, dark-subtracted RGB.** We next took the resulting block-averaged RGB images and subtracted the dark response, characteristic for each color channel and for the image exposure duration, using the dark response values shown in Fig. 5A. After subtracting the dark response, each image was multiplied by the scale factor, to yield a RGB dark-subtracted image standardized to one second exposure duration, ISO 1000 setting, and f-value of 1.8, as prescribed by Eq (1).
4. **Standardized, dark-subtracted RGB to LMS and Luminance formats.** By regressing R,G,B camera sensitivities (measured in Fig. 9) against known L,M,S cone sensitivities, we obtain a  $3 \times 3$  matrix that approximately transforms from the RGB into the LMS color coordinate system. Each pixel of the image can thus be transformed into the LMS system. By multiplying the image in the LMS system with the L,M,S isomerisation rate factors (see Colorimetry above), each image is now expressed in units of L,M,S isomerizations per second and saved as a Matlab  $1007 \times 1519 \times 3$  matrix with name pattern `DSC_####_LMS.mat`. In parallel, the image in the LMS system can be transformed into a grayscale image with pixels in units of  $\text{cd}/\text{m}^2$ , by summing over the three (L, M, S) color channels of each pixel with appropriate weights (see Colorimetry above). The grayscale image is saved as a  $1007 \times 1519$  matrix with name pattern `DSC_####_LUM.mat`.
5. **Image metadata in AUX files.** For each image, the database contains a small auxiliary Matlab structure saved as `DSC_####_AUX.mat` that contains image meta-data: **(i)** the EXIF fields extracted from the JPG and NEF images, among others, the aperture, exposure settings, ISO sensitivity, camera serial number, and timestamp of the image; **(ii)** the identifier of the image in the master database (the image name and album), **(iii)** the timestamp when the image processing was done, **(iv)** fraction of pixels at saturation in the RGB image, **(v)** a warning flag indicating whether the exposure time is out of the linear range or if too many pixels are either at low (dark-response) intensities or at saturation, and **(vi)**, several human-assigned annotations (not available for all images), e.g. the distance to the target, the tripod setting, qualitative time of day, lighting conditions etc. Most images are taken in normal, daylight viewing conditions, and should therefore have no warning flags; some images, however, especially ones that are part of the exposure or time-of-day series, can have saturated or dark pixels, and it is up to the database users to properly handle such images.

## DISCUSSION

We have collected a large variety of images that include many snapshots of what a human observer might conceivably look at, such as images of the horizon and detailed images of the ground, trees, bushes,

and baboons. This riverine / savanna environment in Okavango delta was chosen because it is similar to where human eye is thought to have evolved. The images were taken at various times of the day, and with various distances between the camera and the objects in the scene. A subset of albums focuses on particular objects, such as berries and other edible items, in their natural context. We also provide close-up images of natural objects taken from different distances and accompanied by a ruler (from which an absolute scale of the objects can be inferred). Other albums can be used to explore systematic variation of the image statistics with respect to a controllable parameter, for instance, the variation in the texture of sand viewed from various distances, the change in color composition of the sky viewed at regular time intervals during the day, or the same scene viewed at an increasing angle of the camera from the vertical, from pointing towards the ground to pointing vertically at the sky. A large fraction of the database contains images of various types of Botswana scenery (flood plains, sand plains, woods, grass) sampled without purposely focusing on any particular object.

Many studies of biological visual systems explore the hypothesis that certain measurable properties of neural visual processing systems reflect optimal adaptations to the structure of natural visual environments [10, 11, 24, 47–54]. The images we have collected allow for a reliable estimation of the properties of natural scenes, such as local luminance histograms, spatial two-point and higher-order correlation functions, scale invariance, color and texture content, and the statistics of oriented edges (e.g., colinearity and cocircularity) [55–60] (see also a topical issue on natural systems analysis [2]). Currently, we include only images from a single environment, which allows for adequate sampling of the relevant statistical features that distinguish these natural scenes from random ones. In the future, we plan to expand the dataset to include images of other natural, as well as urban, environments, some of which have already been acquired. This expanded image set will allow for sampling of the statistical features that distinguish different environments.

### Acknowledgements

We thank E. Kanematsu for helpful discussions. This research was supported by grant NEI R01 EY10016 (DHB) and by grants NSF IBN-0344678 (VB), NSF EF-0928048 (VB), NIH R01 EY08124 (VB, PS).

- 
- [1] E. P. Simoncelli and B. A. Olshausen, *Annu Rev Neurosci* **24**, 1193 (2001).
  - [2] W. S. Geisler and D. Ringach, *Vis Neurosci* **26**, 1 (2009).
  - [3] A. Olmos and F. Kingdom, *Mcgill calibrated colour image database* (2004), URL <http://tabby.vision.mcgill.ca>.
  - [4] J. van Hateren, *Nature* **360**, 68 (1992).
  - [5] C. Parraga, J. Vazquez-Corral, and M. Vanrell, *Perception* **36(Suppl)**, 180 (????).
  - [6] J. Atick and A. Redlich, *Neural Computation* **4**, 196 (1992).
  - [7] D. J. Field, *Journal of the Optical Society of America A, Optics and image science* **4**, 2379 (1987).
  - [8] E. Doi, D. Balcan, and M. Lewicki, *IEEE Transactions on Image Processing* **16**, 442 (2007).
  - [9] Y. Karklin and M. Lewicki, *Nature* **457**, 38 (2009).
  - [10] R. C.P., B. Borghuis, Y. Kao, S. P., and V. Balasubramanian, *Proc Natl Acad Sci U S A* **107**, 17368 (2010).
  - [11] P. Garrigan, C. Ratliff, J. Klein, P. Sterling, D. Brainard, and V. Balasubramanian, *PLoS Comput Biol* p. e1000677 (2010).
  - [12] *DARPA image understanding workshop*, URL <http://openlibrary.org/b/OL17266910M/>.
  - [13] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach* (Prentice Hall (NJ, USA), 2002), ISBN 0130851981.
  - [14] A. Torralba and A. Oliva, *Network: Computation in Neural Systems* **14**, 391 (2003).



- [15] W. B. Pennebaker and J. L. Mitchell, *The JPEG Still Image Data Compression Standard* (Van Nostrand Reinhold (Norwell, MA, USA), 1993).
- [16] J. Portilla and E. P. Simoncelli, *International J of Comp Vis* **40**, 49 (2000).
- [17] W. Richards, *Applied Optics* **21**, 2569 (1982).
- [18] G. J. Burton and I. Moorhead, *Applied Optics* **26**, 157 (1987).
- [19] B. A. Olshausen and D. J. Field, *Nature* **381**, 607 (1996).
- [20] A. J. Bell and T. J. Sejnowski, *Vision Research* **37**, 3327 (1997).
- [21] H. Vorhees and T. Poggio, *Nature* **333**, 364 (1998).
- [22] M. Riesenhuber and T. Poggio, *Nature Neuroscience* **3**, 1199 (2000).
- [23] W. S. Geisler, J. Perry, B. Super, and D. Gallogly, *Vision Research* **41**, 711 (2001).
- [24] G. Tkačik, J. Prentice, J. Victor, and V. Balasubramanian, *Proc Natl Acad Sci U S A* (2010).
- [25] D. Kersten, *J Opt Soc Am A* **4**, 2395 (1987), URL [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=3430226%20](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3430226%20).
- [26] D. Chandler and D. Field, *J Opt Soc Am A Opt Image Sci Vis* **24**, 922 (2007), URL [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17361279%20](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17361279%20).
- [27] G. J. Stephens, T. Mora, G. Tkačik, and W. Bialek, *arXiv p. q.bio/0806.2694* (2008).
- [28] A. Stockman and L. Sharpe, *Vision Research* **40**, 1711 (2000).
- [29] *CIE 2007 fundamental chromaticity diagram with physiological axes, Parts 1 and 2. Technical report 170-1. Vienna: Central Bureau of the Commission Internationale de l'Eclairage.*
- [30] L. Yin, R. G. Smith, P. Sterling, and D. H. Brainard, *J Neurosci* **26**, 12351 (2006).
- [31] A. Lee, D. Mumford, and J. Huang, *Intl J of Comp Vis* **41**, 35 (2001).
- [32] *Public domain internet image gallery software*, URL <http://gallery.menalto.com>.
- [33] *Public domain implementation of wGET*, URL <http://www.gnu.org/software/wget/>.
- [34] *Creative commons license*, URL <http://creativecommons.org/licenses/by-nc/3.0/>.
- [35] O. Packer and D. R. Williams, *Light, the retinal image, and photoreceptors. In The Science of Color, 2nd ed, S.K.Shevell, Ed.* (Optical Society of America, Elsevier (Oxford, UK), 2003).
- [36] *Nikon documentation*, URL <http://www.nikonusa.com/template.phpcat=1&grp=2&productNr=25214>.
- [37] URL <http://www.majid.info/mylos/weblog/2004/05/02-1.html>.
- [38] *Public domain image extraction software DCRAW*, URL <http://www.cybercom.net/~dcoffin/dcraw/>.
- [39] *Kodak carousel 4400 projector*, URL <http://slideprojector.kodak.com/carousel/4400.shtml>.
- [40] *Photo research pr-650 spectrascan colorimeter*, URL <http://www.photoresearch.com/current/pr650.aspl>.
- [41] L. Sharpe, A. Stockman, W. Jagla, and H. Jagle, *Journal of Vision* **5**, 948 (2005).
- [42] B. K. P. Horn, *Computer Vision, Graphics and Image Processing* **26**, 135 (1984).
- [43] B. A. Wandell, *IEEE Trans Pattern Analysis and Machine Intelligence* **9**, 2 (1987).
- [44] G. D. Finlayson and M. S. Drew, *Proceedings of 12th IST/SID Color Imaging Conference* pp. 202–204 (1996).
- [45] X. Zhang and D. H. Brainard, *Proceedings of 12th IST/SID Color Imaging Conference (Scottsdale, AZ)* pp. 308–314 (2004).
- [46] D. Brainard, *Spatial Vision* **10**, 433 (1997).
- [47] H. Barlow, *Vision Research* **22**, 635 (1982).
- [48] J. Atick and A. Redlich, *Neural Computation* **4**, 449 (1992).
- [49] L. Maloney, *Journal of the Optical Society of America A* **3**, 1673 (1986).
- [50] B. Regan, C. Julliot, B. Simmen, F. Vienot, P. Charles-Dominique, and J. Mollon, *Philosophical Transaction of the Royal Society of London, B* **356**, 229 (2001).
- [51] J. Lythgoe and J. Partridge, *Journal of Experimental Biology* **146**, 1 (1989).
- [52] L. Chittka and R. Menzel, *Journal of Comparative Physiology A* **171**, 171 (1992).
- [53] D. Osorio and M. Vorobyev, *Proceedings of the Royal Society of London, B* **263**, 593 (1996).
- [54] A. Lewis and L. Zhaoping, *Journal of Vision* **6**, 285 (2006).
- [55] D. Ruderman and W. Bialek, *Phys Rev Lett* **73**, 814 (1994), journal article, URL [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=10057546%20](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10057546%20).
- [56] D. Ruderman, *Network: computation in neural systems* **5**, 517 (1994).
- [57] D. L. Ruderman, *Vision Research* **37**, 3385 (1997).
- [58] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini, *Nat Neurosci* **8**, 1690 (2005).
- [59] M. Sigman, G. Cecchi, C. Gilbert, and M. Magnasco, *Proc Natl Acad Sci U S A* **98**, 1935 (2001), URL [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=11172054%20](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11172054%20).
- [60] Y. Karklin and M. S. Lewicki, *Network* **14**, 483 (2003), URL <http://www.ncbi.nlm.nih.gov/sites/entrez?>

Db=pubmed&Cmd=Retrieve&list\_uids=12938768&dopt=abstractplus.

- [61] R. W. Rodieck, *The First Steps of Seeing* (Sinauer Associates (Sunderland, MA, USA), 1998).
- [62] A. Stockman, L. Sharpe, and C. Fach, *Vision Research* **40**, 2901 (1999).
- [63] R. Bone, J. Landrum, and A. Cains, *Vision Research* **32**, 105 (1992).
- [64] J. Pokorny and V. C. Smith, *The Verriest Lecture. How much light reaches the retina? In C.R. Cavonius, Ed., Colour Vision Deficiencies XIII. Documenta Ophthalmologica Proceedings*, vol. 59 (1997).
- [65] We used version *v5.71* of DCRAW; we note that the output image representation produced by this software is highly version dependent.

**TABLE I. Albums 01a–61a of the Botswana dataset.**

Album	Keywords	Tags
cd01a	baboons, grass, tress, bushes	
cd02a	baboons, grass, tress, bushes	
cd03a	trees	sequence: vertical angle
cd04a	trees, grass	sequence: aperture
cd05a	trees, leaves	sequence: aperture
cd06a	flood plain, grass, trees, sky	
cd07a	baboons, horizon	sequence: aperture
cd08a	sand plain, horizon	sequence: aperture
cd09a	sand plain, bushes, horizon	
cd10a	bushes, trees	sequence: aperture
cd11a	forrest	sequence: aperture
cd12a	forrest, ground	
cd13a	flood plain, water, horizon	
cd14a	horizon, sunset	sequence: time, short+long exposure
cd15a	leaves, grass, ground	
cd16a	road, cars, buildings, town	
cd17a	road, cars, buildings, town	
cd18a	tree, ground	sequence: aperture
cd19a	sand plain, ground	closeup
cd20a	sand plain, horizon	sequence: aperture
cd21a	sand plain, sand ground, cloudy sky	
cd22a	baboons, sand ground, horizon	sequence: aperture
cd23a	leaves, bushes, ground, sky	
cd24a	leaves	sequence: aperture
cd25a	plant, leaves	closeup
cd26a	leaves, bushes	sequence: aperture
cd27a	flood plain, grass, water, horizon	sequence: time, short+long exposure
cd28a	termite mound, horizon	sequence: aperture
cd29a	sand plain, horizon	sequence: aperture
cd30a	baboons, sand plain, bushes, trees, horizon	
cd31a	bushes, termite mound	sequence: scale, sequence: vertical angle
cd32a	flood plain, water, grass, horizon	
cd33a	flood plain, water, grass, horizon	
cd34a	forrest, tree, leaves	sequence: scale
cd35a	grass, trees	sequence: scale
cd36a	bark	closeup, sequence: aperture
cd37a	leaves	sequence: aperture
cd38a	sand plain, trees, horizon, ground	
cd39a	sand plain, horizon	sequence: aperture
cd40a	bark	closeup
cd41a	forrest, trees, leaves, ground	
cd42a	fruit, nuts	closeup, on table
cd43a	grass plain, baboons, water,	
cd44a	baboons, grass, human	
cd45a	sand plain, rock	sequence: scale
cd46a	plain, tree	sequence: scale
cd47a	tree trunk, forrest	sequence: scale
cd48a	tree trunk, sand plain	sequence: scale
cd49a	tree stump, ground	sequence: scale
cd50a	fruit, plant	closeup, sequence: aperture
cd51a	tree, sky	sequence: aperture
cd52a	palm tree	sequence: aperture
cd53a	trees, grass, forest	sequence: time (all day)
cd54a	sand plain, horizon	sequence: time (all day)

**TABLE II. Albums 01b–35b of the Botswana dataset.**

Album	Keywords	Tags
cd01b	dirt, ground	closeup, sequencescale: scale, ruler
cd02b	sand, ground	closeup, sequence: scale, ruler
cd03b	salt deposits, ground	closeup, sequence: scale, ruler
cd04b	scrub, ground	closeup, sequence: scale, ruler
cd05b	sticky grass	closeup, sequence: scale, ruler
cd06b	marula nut	closeup, sequence: scale, ruler
cd07b	sausage fruit	closeup, sequence: scale, ruler
cd08b	elephant dung	closeup, sequence: scale, ruler
cd09b	old figs	closeup, sequence: scale, ruler
cd10b	fresh figs	closeup, sequence: scale, ruler
cd11b	old jackelberry	closeup, sequence: scale, ruler
cd12b	woods, ground	closeup, sequence: scale, ruler
cd13b	fresh buffalo dung	closeup, sequence: scale, ruler
cd14b	fresh jackelberry	closeup, sequence: scale, ruler
cd15b	semiold palm nut	closeup, sequence: scale, ruler
cd16b	old palm nut	closeup, sequence: scale, ruler
cd17b	fresh palm nut	closeup, sequence: scale, ruler
cd18b	fresh sausage fruit	closeup, sequence: scale, ruler
cd19b	semifresh sausage fruit	closeup, sequence: scale, ruler
cd20b	old buffalo dung	closeup, sequence: scale, ruler
cd21b1	marula tree bark	closeup, sequence: scale, ruler
cd21b2	palm tree bark	closeup, sequence: scale, ruler
cd22b1	fig tree bark	closeup, sequence: scale, ruler
cd22b2	jackelberry tree bark	closeup, sequence: scale, ruler
cd23b	sausage tree bark	closeup, sequence: scale, ruler
cd24b1	sage	closeup, sequence: scale, ruler
cd24b2	termite mound	closeup, sequence: scale, ruler
cd25b	woods, horizon, sky	sequence: vertical angle
cd26b	sand plain, horizon, sky	sequence: vertical angle
cd27b	flood plain, water, horizon, sky	sequence: vertical angle
cd28b	bush, sky	sequence: vertical angle
cd29b1	sand plain	sequence: height
cd29b2	woods	sequence: height
cd30b1	flood plain	sequence: height
cd30b2	bush	sequence: height
cd31b	fresh elephant dung	closeup, sequence: scale, ruler
cd32b	sky, no clouds	sequence: time (all day)
cd33b1	horizon	sequence: time (sunrise)
cd33b2	horizon	sequence: time (sunset)
cd34b1	woods, bushes	sequence: time (sunset)
cd34b2	woods, bushes	sequence: time (sunrise)
cd35b	grass, horizon	sequence: time (all day)

Albums with the “ruler” tag have a green ruler present in the scene so that the absolute size of the objects can be determined.

**TABLE III. Cone Parameters for LMS Images**

Cone Parameter	Value	Source
Outer Segment Length	$33\mu m$	[61] Appendix B
Inner Segment Length	$2.3\mu m$	[61] Appendix B
Specific Density	0.5 (axial optical density)	[61] Appendix B
Lens Transmittance	See CVRL database	[62]
Macular Transmittance	See CVRL database	[63]
Pupil Diameter	See [64], Eq.1	[64]
Eye Length	$16.1mm$	[61] Appendix B
Photoreceptor Nomogram	See CVRL database	[28]
Photoreceptor Quantal Efficiency	0.667	[61] page 472

CVRL database is accessible at <http://www-cvrl.ucsd.edu/>.